

基于模态可靠性与可形变跨尺度共注意力的多模态脑肿瘤MRI分割

索国栋^{1,2}, 周月娥^{1,2}, 金丰庆^{1,2}, 杨建兰²

1. 甘肃中医药大学医学信息工程学院, 甘肃 兰州 730000; 2. 泉州市正骨医院, 福建 泉州 362000

【摘要】目的:在 Medical Imaging Segmentation Toolkit 集成的 U-Net 骨干基础上提出一种基于模态可靠性与可形变跨尺度共注意力的三维(3D)多模态脑肿瘤分割网络。**方法:**设计3项轻量级改进:在输入端引入模态可靠性感知融合模块,自适应抑制低质量模态噪声;在解码阶段构建可形变跨尺度共注意力模块,增强肿瘤边界及形变区域的表达;在深监督分支中加入不确定性引导深监督模块,提升对小病灶和模糊区域的优化能力。**结果:**在 BraTS2021 数据集上进行验证,与多种3D分割基线模型进行比较。结果发现本文模型在完整肿瘤、肿瘤核心和增强肿瘤区域均取得性能领先的效果:完整肿瘤区的 Dice 相似性系数(DSC)为 0.921,肿瘤核心区的 DSC 为 0.895,增强肿瘤区的 DSC 为 0.891,平均 95% Hausdorff 距离为 4.122 mm,优于其他模型。**结论:**本研究提出的 3D 多模态脑肿瘤分割网络在保持网络结构轻量化的前提下,有效提升多模态信息融合质量、跨尺度特征对齐能力及深监督优化效率,为脑肿瘤 MRI 自动分割提供一种具有鲁棒性和精细度的方案。

【关键词】脑肿瘤;图像分割;模态可靠性感知融合;跨尺度可形变共注意力;不确定性引导深监督

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2026)07-0930-09

Multi-modal brain tumor MRI segmentation based on modality reliability and deformable cross-scale co-attention

SUO Guodong^{1,2}, ZHOU Yue'e^{1,2}, JIN Fengqing^{1,2}, YANG Jianlan²

1. School of Medical Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, China; 2. Quanzhou Orthopedic-Traumatological Hospital, Quanzhou 362000, China

Abstract: Objective To propose a three-dimensional (3D) multi-modal brain tumor segmentation network built upon the U-Net backbone implemented in the Medical Imaging Segmentation Toolkit, which integrates modality reliability and deformable cross-scale co-attention. **Methods** Three lightweight improvements were designed for the network. Specifically, a modality reliability-aware fusion module was embedded at the input stage to adaptively suppress low-quality modality noise. A deformable cross-scale co-attention module was introduced within the decoder to enhance the representation of tumor boundaries and structurally deformed regions. An uncertainty-guided deep supervision module was used in the deep supervision branch to improve the optimization performance for small lesions and ambiguous regions. **Results** The proposed method was validated on the BraTS2021 dataset and compared with several 3D segmentation baseline models. Results showed that the proposed model achieved optimal performance across whole tumor, tumor core, and enhancing tumor subregions, with Dice similarity coefficients of 0.921, 0.895, and 0.891, respectively, and a mean 95% Hausdorff distance reaching 4.122 mm, which outperformed the other models. **Conclusion** The proposed 3D multi-modal brain tumor segmentation network effectively enhances multi-modal feature fusion, cross-scale feature alignment, and deep supervision optimization, while maintaining a lightweight architecture. This study provides a robust and fine-grained solution for automated brain tumor segmentation from MRI images.

Keywords: brain tumor; image segmentation; modality reliability-aware fusion; deformable cross-scale co-attention; uncertainty-guided deep supervision

【收稿日期】2026-03-21

【基金项目】泉州市科技计划项目(2026QZNY048)

【作者简介】索国栋,硕士研究生,研究方向:医学图像分割与深度学习, E-mail: liupizhi666@163.com

【通信作者】杨建兰,副教授,研究方向:医学影像识别与应用、卫生信息管理系统开发与研究、卫生信息数据挖掘, E-mail: fjjyl@gszy.edu.cn

前言

脑肿瘤是中枢神经系统中高度异质性的恶性疾病,其临床治疗和预后评估高度依赖影像学检查,特别是磁共振成像(Magnetic Resonance Imaging, MRI)对肿瘤核心区(Tumor Core, TC)、肿瘤水肿区(Peritumoral Edema, ED)及增强肿瘤区(Enhancing Tumor, ET)的准确识别具有不可替代的作用^[1-2]。然而,由于人工分割费时费力且依赖专家经验,开发自动化、高精度的脑肿瘤分割方法已成为医学图像分析领域的重要研究方向。医学图像分割任务本身具有结构复杂、病灶形态可变、边界模糊及类别不均衡等难点,特别是在三维(Three-Dimensional, 3D)体数据中,网络需要同时建模3D空间结构关系并保持计算的稳定性^[3-5]。近年来深度学习方法的快速发展显著提高脑肿瘤自动分割的精度,但仍存在挑战,包括多模态MRI之间信噪比差异大、空间刚性或非刚性偏移、模态缺失、成像伪影及肿瘤边界局部对比度不足等,这些因素均可能显著影响模型的稳健性与泛化能力^[6-8]。

为应对上述问题,研究者提出大量基于深度学习的3D卷积神经网络(3D Convolutional Neural Network, 3D CNN)方法^[9-12]。3D U-Net^[10]通过编解码结构与跳跃连接,将局部卷积特征与高分辨率细节进行融合,在医学图像分割中得到广泛应用。V-Net^[11]引入残差结构与Dice损失,在3D医学图像的端到端训练中取得良好表现;nnU-Net^[13]则通过自适应确定网络结构、数据预处理、patch尺寸等策略,成为医学图像分割领域的自动化典型标准。尽管这些模型在多项医学分割任务中表现出色,但它们主要依赖局部卷积操作,缺乏对模态之间关系的显式建模能力,也难以充分利用跨尺度的结构一致性。

为进一步提升多模态MRI的利用效率,许多研究提出多模态融合机制,包括多分支特征提取、模态注意力、模态自适应加权、模态可靠性估计以及跨模态一致性学习等策略。例如,Sun等^[14]通过多条路径分别提取不同特征流,将这些特征在空间层面进行融合,以利用多模态互补信息;Zhuang等^[15]提出跨模态特征交互模块,利用跨模态分支注意力机制,动态地校准不同模态分支的重要性,使得网络在多模态特征的融合过程中能更精准地提取和整合有用的信息;Dolz等^[16]在不同模态的路径之间实现密集连接,能在各个抽象层次内和层次之间学习模态的复杂组合关系。这些方法在模态选择性方面有所提升,但仍难以处理由于成像条件差异导致的空间几何偏移,也鲜少同时建模跨尺度语义一致性。

随着Transformer架构在计算机视觉领域的快速发展,其在多模态脑肿瘤分割中的应用亦大量涌现^[17-19]。例如,TransBTS^[20]将3D CNN与Transformer结合,在编码器中利用多头自注意力建模长程特征依赖,显著提升全局建模能力;Swin-UNETR^[21]利用Swin Transformer^[19]的窗口自注意力与层次化结构,在BraTS^[22-23]等大型数据集上取得优异性能;UNETR++^[24]通过轻量级Transformer编码器与增强的跨层连接机制,实现更高效的跨尺度信息流动。此外,nnFormer^[25]通过局部窗口注意力与跨窗口全局注意力相结合,在3D肿瘤分割中展现出更强的结构建模能力。尽管这些方法有效增强全局依赖建模,但普遍存在参数量大、训练成本高及对数据规模敏感等问题。

除注意力机制之外,结构对齐与跨尺度几何一致性建模也成为近年来的研究热点。例如,基于可形变卷积或空间变换网络的几何对齐方法能在一定程度上解决不同尺度、不同模态之间的空间不一致问题^[26];多尺度可形变注意力模型则进一步增强局部结构和全局语义的协同推理能力^[27-28]。同时,卷积网络结构如MedNeXt^[29]通过采用近似Transformer的分层扩张卷积模块,在性能、速度与参数量之间实现更优平衡。

综上,当前多模态脑肿瘤分割面临的主要问题包括:(1)模态质量不一致导致的融合不稳定;(2)跨尺度特征之间存在几何偏移,却缺乏有效的对齐策略;(3)深监督分支在复杂结构学习中存在梯度不稳定与噪声敏感问题。针对上述挑战,本研究在经典U-Net的基础上提出一种轻量化、可插拔、训练友好的多模态分割新框架,包含以下核心创新:(1)模态可靠性感知融合(Modality Reliability-Aware Fusion, MRAF),针对多模态输入质量不一致问题,自适应估计每个模态的像素级可靠性权重,从而提升融合有效性;(2)跨尺度可形变共注意力(Deformable Cross-Scale Co-Attention, DCSCA),在解码阶段对跳跃连接进行空间可形变对齐与跨尺度语义建模,提高结构一致性与边界区域的精细分割能力;(3)不确定性引导深监督(Uncertainty-Guided Deep Supervision, UGDS),通过空间不确定性图对深监督输出进行动态调整,提高网络在训练初期与收敛阶段的稳定性。实验结果显示,该方法在BraTS2021^[30]数据集上优于包括3D U-Net、nnU-Net、TransBTS、Swin-UNETR、UNETR++、MedNeXt的多种经典与最新模型,展现出良好的性能与泛化能力。

1 方法

1.1 网络总体结构

本研究在 Medical Imaging Segmentation Toolkit (MIST)^[31-32] 集成的 U-Net 模型基础上,设计一种面向

多模态脑肿瘤 MRI 的轻量化改进网络。整体结构如图 1 所示,网络以 4 通道多模态 MRI 体数据 (T_1 、 T_1Gd 、 T_2 、FLAIR) 作为输入,经过 MRAF 模块对各模态进行像素级置信建模与加权,再送入 3D 编码解码主干网络。

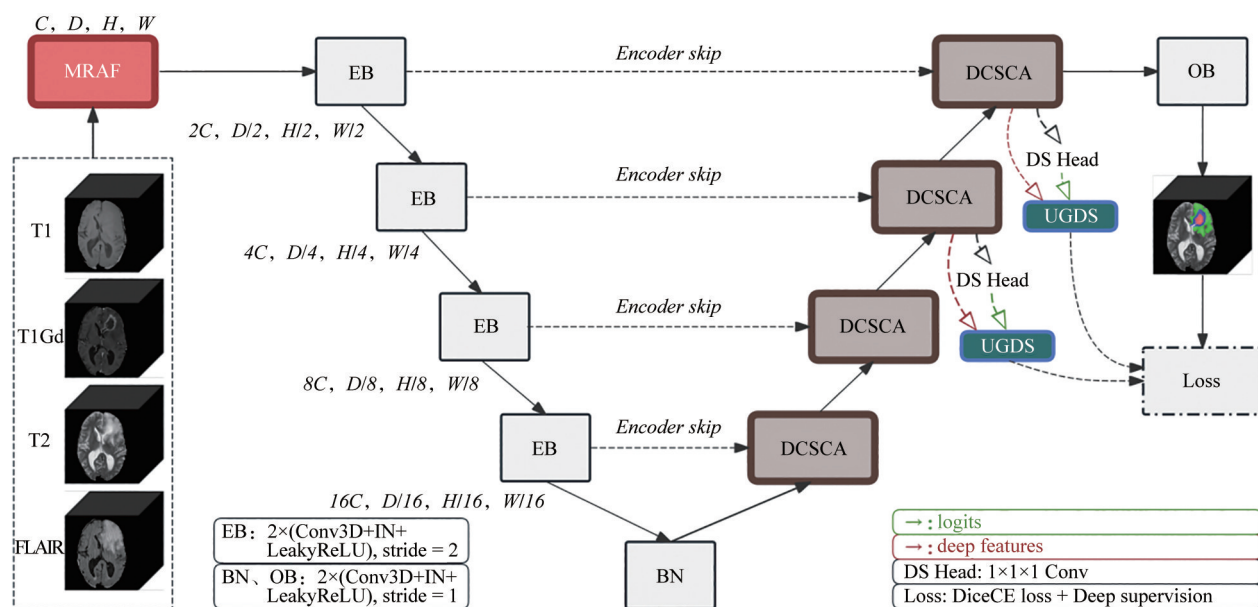


图1 整体网络结构图

Figure 1 Overall architecture of the proposed network

编码器部分由多级 3D 卷积块与步长下采样组成,用于逐步提取从浅层纹理到高层语义的多尺度体积特征;解码器通过上采样操作逐级恢复空间分辨率,并与对应尺度的编码器特征通过跳跃连接进行融合,以保留关键的结构细节。在此基础上,本研究在 3 个关键位置引入轻量化“插拔式”模块:在进入主干网络前使用 MRAF 对多模态进行可靠性加权融合,抑制低质量模态的干扰;在每一级解码前,对当前解码特征与对应编码器特征引入 DCSCA 模块,实现几何对齐与语义协同;在原有多尺度深监督输出头之后引入 UGDS 模块,对深监督预测进行空间自适应重加权。需要强调的是,本研究并未改动主干的卷积结构和原有损失函数形式,而是通过这 3 种轻量化模块对多模态可靠性建模、跨尺度对齐和深监督稳定性进行增强,从而在保持网络复杂度可控的前提下提升分割性能与鲁棒性。

1.2 MRAF 模块

多模态脑肿瘤 MRI 在不同序列和不同病例之间的信噪比、对比度和伪影分布存在明显差异,简单地将各模态在通道维度等权拼接,容易将低质量模态噪声一并引入网络,影响肿瘤边界和亚区分割的精

度。为此,本研究在网络最前端提出 MRAF 模块,对各模态建立像素级置信权重,并通过残差门控的方式实现稳健融合,如图 2 所示。

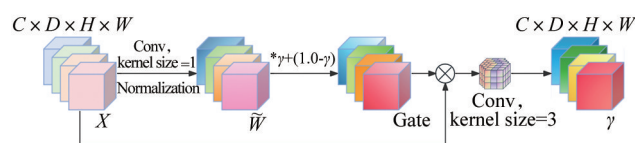


图2 MRAF 模块

Figure 2 MRAF module

设输入体数据为 $X \in \mathbb{R}^{B \times C \times D \times H \times W}$, 其中, B 为 batch 大小, C 为通道数, D 、 H 、 W 分别为空间维度。MRAF 首先使用一个 $1 \times 1 \times 1$ 卷积生成每个模态的响应图:

$$L = f_{\text{conv}}(X), L \in \mathbb{R}^{B \times C \times D \times H \times W} \quad (1)$$

随后对 L 采用 Softmax 归一化,得到像素级模态权重 W :

$$W = \text{Softmax}\left(\frac{L}{\tau}\right)_{\text{channel}} \quad (2)$$

其中, τ 为温度系数,用于控制权重分布的锐利程度。

为避免某模态权重被压缩到0, 进一步设置下限 ε 并重新归一化:

$$\tilde{W} = \frac{\max(W, \varepsilon)}{\sum_{c=1}^C \max(W_c, \varepsilon)} \quad (3)$$

在得到权重后, MRAF 采用残差门控的形式进行融合:

$$G = \gamma \tilde{W} + (1 - \gamma), Y = X \otimes G \quad (4)$$

其中, γ 为门控强度, 当 $\gamma=0$ 时退化为恒等映射, 当 $\gamma>0$ 时逐渐强化模态加权效果。为抑制孤立噪声响应, MRAF 通过一个深度可分离 $3 \times 3 \times 3$ 卷积对 Y 进行轻量平滑。最终, 经 MRAF 得到的 Y 替代原始 X 作为主

干的输入, 在不引入额外监督的前提下显式建模模态质量差异。

1.3 DCSCA 模块

传统 U-Net 结构中, 解码器往往通过简单的拼接或相加操作与对应尺度的编码器特征进行融合, 未显式考虑二者在空间对齐和语义层次上的差异。对于具有复杂形变和模态差异的脑肿瘤 MRI, 这种朴素的跳跃连接容易导致结构信息对齐不足, 影响小病灶和边界区域的分割精度。为此, 本研究在每一层解码前引入 DCSCA 模块, 对当前解码特征与编码器特征进行几何对齐和语义协同建模, 细节实现如图3所示。

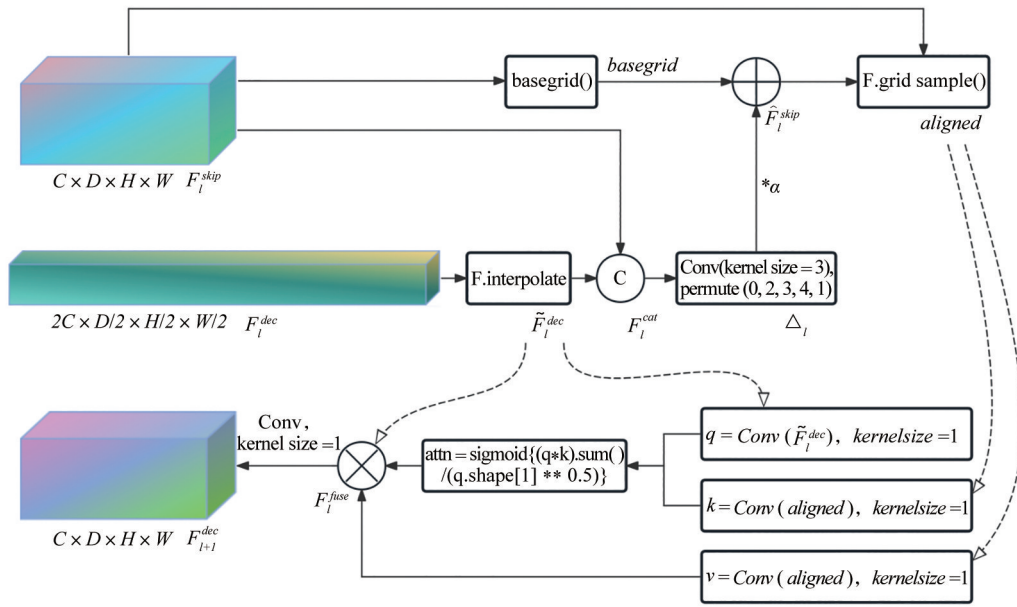


图3 DCSCA 模块

Figure 3 DCSCA module

设第 l 层解码器特征为 F_l^{dec} , 对应的编码器特征为 F_l^{skip} 。首先, 将 F_l^{dec} 插值到与 F_l^{skip} 相同的空间分辨率, 得到 $\tilde{F}_l^{dec}$:

$$\tilde{F}_l^{dec} = \text{Interp}(F_l^{dec}, \text{size}(F_l^{skip})) \quad (5)$$

然后, 将两者在通道维度拼接:

$$F_l^{cat} = \text{Concat}(\tilde{F}_l^{dec}, F_l^{skip}) \quad (6)$$

通过一个 $3 \times 3 \times 3$ 卷积预测位移场, 在归一化坐标系下加到基础网格 G_0 上, 并利用 GridSample() 对编码器特征进行可形变采样, 得到几何对齐后的特征:

$$\hat{F}_l^{skip} = \text{GridSample}(F_l^{skip}, G_0 + \alpha \Delta_l) \quad (7)$$

其中, α 为控制形变幅度的缩放系数。在几何对齐之后, DCSCA 进一步在高分辨率特征上构建跨尺度共注意力。通过 $1 \times 1 \times 1$ 卷积分别得到查询 Q_l 、键 K_l 与值 V_l :

$$Q_l = f_Q(\tilde{F}_l^{dec}), K_l = f_K(\hat{F}_l^{skip}), V_l = f_V(\hat{F}_l^{skip}) \quad (8)$$

采用通道点乘的方式得到空间注意力图:

$$A_l = \sigma\left(\frac{\sum_c Q_{l,c} \cdot K_{l,c}}{\sqrt{C}}\right) \quad (9)$$

其中, C 为嵌入通道数, $\sigma()$ 为 Sigmoid 函数。最终的融合特征为:

$$F_l^{fuse} = V_l \otimes A_l + \tilde{F}_l^{dec} \quad (10)$$

通过一个 $1 \times 1 \times 1$ 卷积映射回与 F_l^{skip} 相同的通道数, 作为更新后的 skip 特征输入到后续的上采样模块中。

1.4 UGDS 模块

主干在多个尺度上引入深度监督输出, 以缓解梯度消失并加速收敛。在训练初期或者标注边界模

糊区域,深监督分支的预测往往存在较大不确定性,若直接与主分支等权计算损失,可能会引入噪声梯度,反而影响整体性能。为此,本研究提出UGDS模块,在不改变原有损失函数形式的前提下,对多尺度深监督预测进行空间自适应重加权,如图4所示。

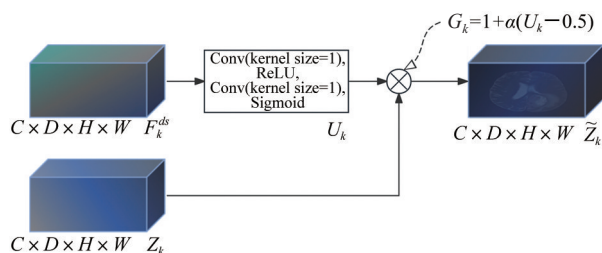


图4 UGDS模块

Figure 4 UGDS module

设第 k 个深监督层的特征图为 F_k^{ds} ,原始深监督 logits 为 Z_k 。UGDS首先从 F_k^{ds} 预测1个单通道不确定性图 $U_k \in [0, 1]$:

$$U_k = \sigma(f_{\text{uncertainty}}(F_k^{ds})) \quad (11)$$

其中, $f_{\text{uncertainty}}$ 为两层 $1 \times 1 \times 1$ 卷积和ReLU组合。随后构造空间门控图:

$$G_k = 1 + \alpha(U_k - 0.5) \quad (12)$$

其中, α 为不确定性门控强度超参数,本研究取0.6。最终,对 logits 进行重加权得到修正后的深监督输出:

$$\tilde{Z}_k = Z_k \otimes G_k \quad (13)$$

在损失计算阶段,仍沿用原有框架中的深监督损失形式,仅将 $\tilde{Z}_k$ 作为该尺度的预测输入至损失函数。UGDS不引入额外监督信号,也不改变目标函数定义,仅通过学习到的不确定性图抑制噪声较大的位置,从而使深监督更好地服务于主分支学习。

1.5 损失函数

1.5.1 Dice交叉熵损失 本研究的主分支损失函数沿用常用的Dice交叉熵组合形式,对前景-背景及各种瘤亚区进行联合优化。设预测概率图为 $P \in R^{B \times K \times D \times H \times W}$,对应的 one-hot 形式真实标注为 $Y \in \{0, 1\}^{B \times K \times D \times H \times W}$,其中, K 为类别数。多类软Dice损失可写为:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{b,k} \sum_i P_{b,k,i} Y_{b,k,i} + \mu}{\sum_{b,k} \sum_i P_{b,k,i}^2 + \sum_{b,k} \sum_i Y_{b,k,i}^2 + \mu} \quad (14)$$

其中, i 遍历空间位置, μ 为避免分母为0的平滑项。多类交叉熵损失定义为:

$$L_{\text{CE}} = -\frac{1}{\text{BDHW}} \sum_b \sum_i \sum_k Y_{b,k,i} \log P_{b,k,i} \quad (15)$$

综合考虑区域重叠度与分类概率分布,本研究

采用二者加权和作为总损失:

$$L_{\text{main}} = \lambda_{\text{Dice}} L_{\text{Dice}} + \lambda_{\text{CE}} L_{\text{CE}} \quad (16)$$

其中, λ_{Dice} 与 λ_{CE} 为权重系数,实验中采用 $\lambda_{\text{Dice}} = \lambda_{\text{CE}} = 0.5\lambda$,与原始MIST中集成U-Net框架保持一致。

1.5.2 深度监督策略 为缓解3D分割任务中梯度衰减问题并加速收敛,本研究在多个解码层引入深监督输出,采用和主分支相同的Dice交叉熵组合损失。设共有 L_{ds} 个深监督分支,第 k 个分支的修正后 logits 为 $\tilde{Z}_k$,对应的概率图为 $\tilde{P}_k$ 。在计算损失前,将各尺度预测上采样到与原始标签相同的空间尺寸,再计算其Dice-CE损失:

$$L_{\text{ds}}^{(k)} = \lambda_{\text{Dice}} L_{\text{Dice}}^{(k)} + \lambda_{\text{CE}} L_{\text{CE}}^{(k)} \quad (17)$$

深监督的总损失为各尺度损失的加权和:

$$L_{\text{DS}} = \sum_{k=1}^{L_{\text{ds}}} \beta_k L_{\text{ds}}^{(k)} \quad (18)$$

其中, β_k 为不同尺度的权重系数,通常对高分辨率分支赋予更大权重,对低分辨率分支赋予较小权重。最终的总体训练损失为:

$$L = L_{\text{main}} + L_{\text{DS}} \quad (19)$$

需要指出的是,本研究并未改变原有Dice-CE和深度监督的损失形式,而是在损失计算前通过UGDS对深监督 logits 进行不确定性加权,从而在不增加额外目标项的前提下,提高深监督提供的梯度信息质量。

2 实验

2.1 数据集

本研究使用BraTS2021数据集,包括1251例多模态3D脑肿瘤MRI数据,每例病例均包含 T_1 、 $T_1\text{Gd}$ 、 T_2 、FLAIR这4种标准MRI模态,并配有完整的肿瘤区域标注,包括完整肿瘤区(Whole Tumor, WT)、TC和ET这3个组织结构,任务目标是从四模态输入中自动分割上述肿瘤区域。为保证模型评估的稳定性与泛化能力,数据集首先随机划分出10%作为最终测试集,其余90%用于五折交叉验证训练,其中每折分别构建训练集与验证集并取五折平均性能作为最终结果。所有影像在训练前均经过标准化处理,包括NiftI转NumPy、重采样至统一体素分辨率、z-score强度归一化、前景区域自动裁剪以及生成 $128 \times 128 \times 128$ 的3D patch;同时,采用oversampling确保采样更关注肿瘤区域。训练阶段进一步引入多种随机数据增强,如翻转、模糊、噪声、亮度与对比度扰动、随机缩放等,以提升模型的鲁棒性。

2.2 评价指标

本研究采用Dice相似系数(Dice Similarity Coefficient, DSC)与95% Hausdorff距离(95%

Hausdorff Distance, HD95)作为主要评价指标,以全面衡量模型在脑肿瘤分割任务中的重叠精度与边界逼近程度。对于预测分割结果 P 与真实标注 G ,DSC用于评估两者的重叠程度,定义为:

$$DSC(P,G)=\frac{2|P\cap G|}{|P|+|G|}$$

(20)

其取值范围为 $[0,1]$,值越高表示预测区域与真实肿瘤区域重叠越好。

HD用于度量两组边界点之间的最大偏差,但原始HD对异常噪声较为敏感,因此本研究采用其95%分位数形式获得更稳健的边界差异估计,即:

$$HD95(P,G)=\max\{\text{quantile}_{95}(d(P,G)),\text{quantile}_{95}(d(G,P))\}$$

(21)

其中, $d(P,G)$ 表示从预测边界到真实边界的逐点最小欧氏距离。

DSC反映分割区域的整体准确性,而HD95则关注分割边界的吻合程度,两者结合可全面评价模型在WT、TC、ET这3个肿瘤区域上的分割性能。

2.3 实验环境

本研究在 Ubuntu22.04 操作系统上基于 Python3.10、PyTorch2.1.0 和 CUDA12.1 实现,全部实验均在单张 24 GB 显存的 NVIDIA RTX 4090 GPU 和 16 核 Intel Xeon Gold 6430 vCPU 的计算环境中完成。网络实现基于 MIST 框架中集成的 U-Net 体系,并在其基础上扩展 MRAF、DCSCA 和 UGDS 模块。Epoch 大小为 200,训练中使用自动混合精度以提升效率,并采用 Adam 优化器,学习率初始化为 3×10^{-4} ,配合 cosine 学习率调度策略。

3 结果

3.1 模型对比结果

在 BraTS2021 数据集上,对比 6 种具有代表性的 3D 脑肿瘤分割模型,包括 3D U-Net、nnU-Net、TransBTS、Swin-UNETR、UNETR++ 和 MedNeXt。这些模型在不同肿瘤区域上具有明显的性能差异,其中基于 Transformer 或多尺度编码的架构在 WT 和 TC 的 DSC 上略优于传统 CNN,而 nnU-Net 仍保持强基线性能。然而,本文方法在所有 3 个肿瘤区域均取得最优的 DSC: WT 为 0.921, TC 为 0.895, ET 为 0.891, 平均 DSC 达到 0.902。在 HD95 指标上,本文方法同样表现最好,平均 HD95 为 4.122 mm,显著优于所有对比模型。特别是在较难分割的 ET 区域,本文方法在 DSC 与 HD95 上均表现出明显优势,说明所提出的 MRAF、DCSCA 与 UGDS 机制有效提升模型对小病灶及边界模糊区域的辨识能力。图 5 展示五折交叉验证时的训练与验证损失随迭代轮数变化的趋势,

阴影部分表示每个 epoch 的标准差区间,该损失曲线说明所提出网络结构和优化策略能实现稳定训练,并在不同数据划分下均保持可靠的泛化表现。总体而言,数据结果证实本文方法在 3 类肿瘤区域均取得最稳健、最全面的分割性能,在 DSC 与 HD95 两项指标上均优于现有对比方法。

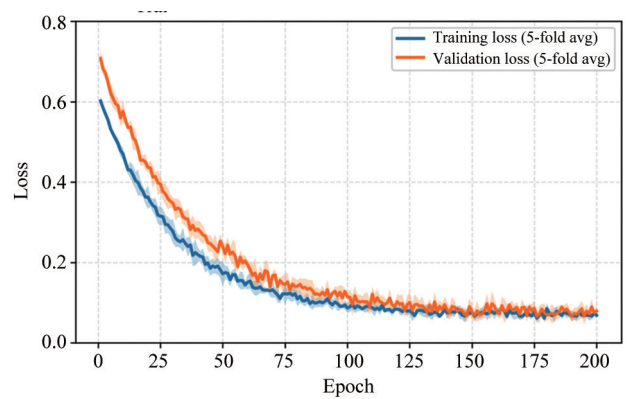


图5 五折交叉验证时的训练与验证损失变化曲线图
Figure 5 Training and validation loss curves under 5-fold cross-validation

图 6 展示在测试集中选取的 3 个典型病例的分割可视化对比结果,依次给出 6 个分割模型、本文方法以及真实标签(Ground Truth, GT)的分割图。绿色、紫色和红色分别表示肿瘤水肿、增强肿瘤以及肿瘤坏死区域,黑色代表背景。从图中可以观察到 3D U-Net、TransBTS 结构在 WT 普遍出现碎块或边缘不完整的问题,尤其在病例 1 的肿瘤水肿和肿瘤坏死区域中遗漏较为明显。nnU-Net、UNETR++ 和 MedNeXt 整体分割质量较高,但在病例 1 的肿瘤坏死区域、病例 2 和病例 3 的肿瘤水肿与增强肿瘤边界存在轻微缺损或不连续现象,且部分模型在病例 3 的水肿区域存在局部过分扩张的情况。本文方法的分割结果整体最接近 GT,在 3 种肿瘤区域上均展现出更高的结构准确性、更少的假阳性碎块,同时对边界较复杂的病灶仍能保持较好的分割一致性,与 GT 的视觉重叠度显著高于对比模型,这与表 1 中量化指标的提

3.2 消融实验

首先对 MRAF 的关键超参数进行小规模参数敏感性测试,结果如表 2 所示。其中, τ 表示温度系数,用于调节模态权重分布的平缓程度; γ 表示残差门控强度系数,用于控制可靠性权重对输入特征的调制程度; ε 表示模态权重下限系数,用于避免部分模态信息被过度抑制;空间平滑处理表示是否利用深度卷积对门控后的特征进行轻量平滑;权重梯度分离表示是否阻断模态权重分支在门控过程中的梯度传

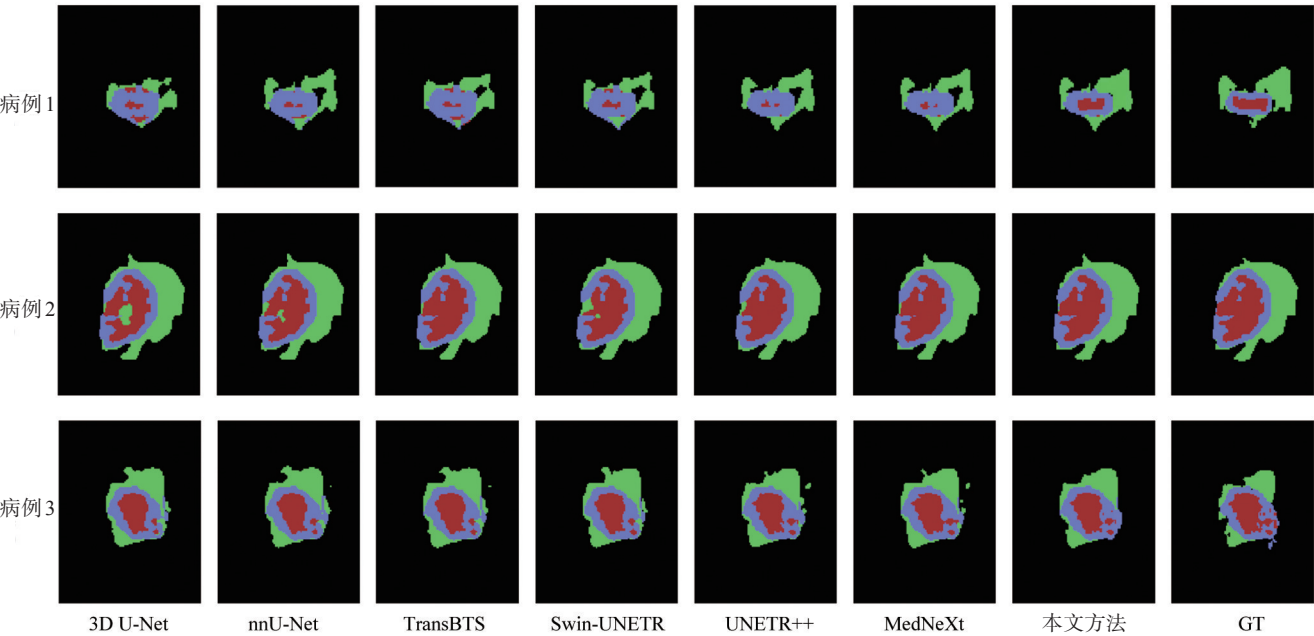


图6 可视化对比图
Figure 6 Visual comparison

表1 在 BraTS2021 数据集上的对比实验结果
Table 1 Comparative results of different models on the BraTS2021 dataset

模型	DSC				HD95/mm			
	WT	TC	ET	平均	WT	TC	ET	平均
U-Net(3D)	0.894	0.839	0.815	0.849	3.126	6.152	6.832	5.370
nnU-Net	0.913	0.871	0.851	0.878	2.757	4.825	5.972	4.518
TransBTS	0.904	0.842	0.852	0.866	3.104	5.901	4.465	4.490
Swin-UNETR	0.919	0.862	0.850	0.877	4.257	4.912	5.916	5.028
UNETR++	0.917	0.876	0.850	0.881	3.561	4.477	5.372	4.470
MedNeXt	0.916	0.860	0.861	0.879	4.169	5.161	5.813	5.048
本文方法	0.921	0.895	0.891	0.902	3.027	4.301	5.038	4.122

播。实验中共评估3组参数组合,其中包含较为保守的弱门控设置以及更激进的强门控设置。结果显示, $\tau=2.0$ 、 $\gamma=0.2$ 、 $\varepsilon=0.10$ 的弱门控在整体性能上,平均DSC为0.893; $\tau=1.0$ 、 $\gamma=0.5$ 、 $\varepsilon=0.02$ 的强门控方案甚至略降至0.889,说明过度平滑或过度集中的模态权重分布均不利于模型的鲁棒表达。相比之下, $\tau=1.5$ 、

$\gamma=0.3$ 、 $\varepsilon=0.05$ 的中等强度门控配置取得最佳表现,平均DSC为0.902。

因此,在后续全部消融实验中,均采用这一组最优参数作为MRAF的默认配置,实验结果如表3所示。总体来看,3个模块均能不同程度提升基线模型的分割精度,其中UGDS对ET与WT的提升更为显

表2 MRAF 模块超参数敏感性分析结果
Table 2 Hyperparameter sensitivity analysis of the MRAF module

温度系数 τ	残差门控强度系数 γ	模态权重下限系数 ε	空间平滑处理	权重梯度分离	平均DSC
2.0	0.2	0.10	√	√	0.893
1.0	0.5	0.02	×	×	0.889
1.5	0.3	0.05	√	×	0.902

著,而MRAF和DCSCA则分别在模态融合与跨尺度特征对齐方面提供稳定增益。单独加入MRAF后,平均DSC由0.874提升至0.884,HD95从4.951 mm降至4.798 mm,表明该模块能有效实现模态可靠性建模并增强输入特征的判别能力。加入DCSCA也带来类似的表现,平均DSC达到0.886,显示出跨尺度可形变对齐与轻量共注意力能进一步减少解码端跳跃连接中的空间错位问题,从而减小结构性误分割。相比之下,单独引入UGDS可以使平均DSC提升至

0.888,同时HD95显著下降至4.749 mm,说明基于不确定性引导的深监督策略能够有效抑制深层预测中的噪声,并强化对肿瘤边界区域的学习。当任意两个模块组合时,性能整体呈现协同增益。例如,MRAF+DCSCA组合将平均DSC提升到0.895,而DCSCA+UGDS的HD95则进一步降至4.344 mm,验证输入端模态表征、解码端跨尺度对齐以及深监督可靠性引导之间具有良好的互补性。

表 3 不同模块策略消融实验
Table 3 Ablation study of different module configurations

模型	模块			DSC				HD95/mm			
	MRAF	DCSCA	UGDS	WT	TC	ET	平均	WT	TC	ET	平均
Baseline				0.898	0.873	0.851	0.874	3.891	5.146	5.816	4.951
Baseline+MRAF	√			0.909	0.879	0.865	0.884	3.754	4.913	5.727	4.798
Baseline+DCSCA		√		0.905	0.880	0.872	0.886	3.761	4.874	5.690	4.775
Baseline+UGDS			√	0.912	0.879	0.873	0.888	3.625	4.919	5.702	4.749
Baseline+MRAF+DCSCA	√	√		0.914	0.889	0.883	0.895	3.282	4.571	5.447	4.433
Baseline+MRAF+UGDS	√		√	0.917	0.883	0.878	0.893	3.577	4.641	5.481	4.566
Baseline+DCSCA+UGDS		√	√	0.914	0.889	0.882	0.895	3.211	4.512	5.309	4.344
本文方法	√	√	√	0.921	0.895	0.891	0.902	3.027	4.301	5.038	4.122

4 结语

本研究围绕多模态脑肿瘤MRI分割中存在的模态质量不一致、跨尺度特征错配及深监督训练不稳定等关键挑战,提出一种基于模态可靠性与可形变跨尺度共注意力的3D轻量化分割框架。通过引入3项具有可插拔特性的改进模块(MRAF、DCSCA以及UGDS),本文方法在不增加过多参数量和计算复杂度的前提下,显著提升多模态特征利用效率、跳跃连接特征几何一致性以及深监督优化的稳定性。在BraTS2021数据集上的实验验证表明,本文模型的平均DSC和平均HD95指标均实现优于对比模型的分割性能,定量结果与可视化分析共同验证方法的有效性与稳定性。其中,MRAF有效提升多模态融合对低质量输入的鲁棒性,DCSCA增强网络在形变及复杂边界区域的表达能力,UGDS则进一步提高深监督分支在小病灶与高不确定性区域的优化能力。总体而言,本研究提出的轻量级多模态脑肿瘤分割框架具有良好的性能、可扩展性与部署潜力,可作为脑肿瘤影像自动分割流程中的高效模块,为提升脑肿瘤放疗计划制定、术前评估等临床任务的自动化水平

提供一种可靠的技术路径。未来工作将进一步探索模态缺失情况下的鲁棒分割策略、更精细的边界建模方法,以推动多模态脑肿瘤自动分割技术向临床应用落地。

【参考文献】

[1] Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary[J]. Acta Neuropathol, 2016, 131(6): 803-820.

[2] van den Bent MJ, Geurts M, French PJ, et al. Primary brain tumours in adults[J]. Lancet, 2023, 402(10412): 1564-1579.

[3] Hesamian MH, Jia WJ, He XJ, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. J Digit Imaging, 2019, 32(4): 582-596.

[4] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis[J]. Med Image Anal, 2017, 42: 60-88.

[5] 刘海超, 宋丽娟. 多模态MRI脑肿瘤分割方法的特征融合技术综述[J]. 计算机工程与应用, 2024, 60(23): 28-48.

[6] Liu HC, Song LJ. Review of feature fusion techniques for multimodal MRI brain tumor segmentation methods[J]. Computer Engineering and Applications, 2024, 60(23): 28-48.

[7] Verma A, Yadav AK. Brain tumor segmentation with deep learning: current approaches and future perspectives[J]. J Neurosci Methods, 2025, 418: 110424.

[7] Zaitoon R, Syed H. A comprehensive review: tumour recognition, segmentation and classification in brain MRI[C]//2024 International Conference on Expert Clouds and Applications (ICOECA). Piscataway, NJ, USA: IEEE, 2024: 865-870.

- [8] 王楠, 王华, 魏德健, 等. U-Net及其变体在脑肿瘤图像分割中的应用进展[J]. 激光与光电子学进展, 2025, 62(16): 14-34.
Wang N, Wang H, Wei DJ, et al. Applications and advancements of U-Net and its variants in brain tumor image segmentation[J]. Laser & Optoelectronics Progress, 2025, 62(16): 14-34.
- [9] Brosch T, Tang LY, Yoo Y, et al. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation[J]. IEEE Trans Med Imaging, 2016, 35(5): 1229-1239.
- [10] Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016. Cham: Springer International Publishing, 2016: 424-432.
- [11] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). Piscataway, NJ, USA: IEEE, 2016: 565-571.
- [12] Kamnitsas K, Ledig C, Newcombe VF, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Med Image Anal, 2017, 36: 61-78.
- [13] Isensee F, Jaeger PF, Kohl SA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nat Methods, 2021, 18(2): 203-211.
- [14] Sun JD, Peng YJ, Guo YF, et al. Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN[J]. Neurocomputing, 2021, 423: 34-45.
- [15] Zhuang YZ, Liu H, Song EM, et al. A 3D cross-modality feature interaction network with volumetric feature alignment for brain tumor and tissue segmentation[J]. IEEE J Biomed Health Inform, 2023, 27 (1): 75-86.
- [16] Dolz J, Gopinath K, Yuan J, et al. HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation[J]. IEEE Trans Med Imaging, 2019, 38(5): 1116-1126.
- [17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [18] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03)[2025-08-20]. <https://arxiv.org/abs/2010.11929>.
- [19] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[EB/OL]. (2021-08-17)[2025-09-15]. <https://arxiv.org/abs/2103.14030>.
- [20] Wang WX, Chen C, Ding M, et al. TransBTS: multimodal brain tumor segmentation using transformer[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 109-119.
- [21] Hatamizadeh A, Nath V, Tang YC, et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images [EB/OL]. (2022-01-4)[2025-09-13]. <https://arxiv.org/abs/2201.01266>.
- [22] Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS)[J]. IEEE Trans Med Imaging, 2015, 34(10): 1993-2024.
- [23] Bakas S, Akbari H, Sotiras A, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features[J]. Sci Data, 2017, 4: 170117.
- [24] Shaker A, Maaz M, Rasheed H, et al. UNETR++: delving into efficient and accurate 3D medical image segmentation[J]. IEEE Trans Med Imaging, 2024, 43(9): 3377-3390.
- [25] Zhou HY, Guo JS, Zhang YH, et al. nnFormer: volumetric medical image segmentation via a 3D transformer[J]. IEEE Trans Image Process, 2023, 32: 4036-4045.
- [26] Dai JF, Qi HZ, Xiong YW, et al. Deformable convolutional networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 764-773.
- [27] Zhu XZ, Su WJ, Lu LW, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. (2021-03-18)[2025-10-20]. <https://arxiv.org/abs/2010.04159>.
- [28] Chen JN, Lu YY, Yu QH, et al. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-8)[2025-09-21]. <https://arxiv.org/abs/2102.04306>.
- [29] Roy S, Koehler G, Ulrich C, et al. MedNeXt: transformer-driven scaling of ConvNets for medical image segmentation[EB/OL]. (2024-06-02)[2025-09-13]. <https://arxiv.org/abs/2303.09975>.
- [30] Baid U, Ghodasara S, Mohan S, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification [EB/OL]. (2021-09-12)[2026-01-10]. <https://arxiv.org/abs/2107.02314>.
- [31] Celaya A, Actor JA, Muthusivarajan R, et al. PocketNet: a smaller neural network for medical image analysis[J]. IEEE Trans Med Imaging, 2023, 42(4): 1172-1184.
- [32] Celaya A, Lim E, Glenn R, et al. MIST: a simple and scalable end-to-end 3D medical imaging segmentation framework[EB/OL]. (2024-11-18)[2025-09-11]. <https://arxiv.org/abs/2407.21343>.

(编辑:谭斯允)